8/PRTS

Our Ref.: B 3270 PCT

Novel means and methods for the preparation and activation of nucleoside and nucleotide based drugs

The present invention relates to novel means and methods for the preparation and activation of nucleoside and nucleotide based drugs. In particular, the present invention involves a method for the production of a polypeptide having or having enhanced kinase activity for a nucleoside or nucleotide analog and to polypeptides obtainable by said method. The present invention also relates to polynucleotides and vectors encoding said polypeptide obtainable by the method of the invention as well as to host cells transformed therewith. Antibodies against said polypeptide are also within the scope of the present invention. The present invention additionally relates to pharmaceutical and diagnostic compositions as well as kits comprising proteins having kinase activity for a nucleoside or nucleotide analog or the before described polypeptides, polynucleotides, vectors and antibodies. Furthermore, the present invention relates to the use of the before described proteins, polypeptides, polynucleotides, vectors and antibodies for the preparation of pharmaceutical compositions for treating, preventing and/or delaying a disease related to viral infection or cancer. In addition, the present invention relates to a method for identifying inhibitors of nucleoside or nucleotide kinases and to methods for identifying nucleoside or nucleotide based prodrugs employing the above mentioned polypeptides, polynucleotides, vectors and host cells. Also the invention relates to the compounds identifiable by said methods as well as to pharmaceutical and diagnostic compositions comprising said inhibitors. Moreover, the present invention relates to the use of proteins and polypeptides having nucleoside or nucleotide kinase activity or their encoding polynucleotides or vectors for the preparation of nucleoside or nucleotide phosphates or analogs and derivatives thereof.

Several documents are cited throughout the text of this specification. Each of the documents cited herein (including any manufacturer's specifications, instructions, etc.) are hereby incorporated herein by reference; however, there is no admission that any document cited is indeed prior art as to the present invention.

The class of medicinal agents termed prodrugs are compounds that exert a desired biological effect only after some sort of modification inside the body. AZT and d4T are examples of nucleoside analog prodrugs that attain their medicinal activity only after being phosphorylated three times to their triphosphate form by cellular kinases. Thus, the efficacy of these compounds is determined in part by the efficiency of the kinases that activate them. Compounds that are poorly activated (as is the case for AZT) do not achieve their full therapeutic potential. Moreover, the intermediate metabolites between the administered prodrug and the active triphosphate form can be toxic. Therefore, it is very important to use prodrugs that are efficiently transformed to their active form or to improve the activity of the rate limiting enzyme(s) in the activation pathway of a prodrug. The latter approach, however, was challenged in the art as being naive and doubtful to work at all; see, e.g., Balzarini, Nature Medicine 4 (1998), 2. Recently, Guettari, (Virology 235 (1997), 398-405) described the improvement of AZT metabolism by use of the Herpes Simplex virus-1 thymilidate kinase (HSV-1 TK) and suggested that gene transfer might be envisioned for genetic pharmacomodulation of anti-viral drugs. However, the extent to which HSV-1 TK improved AZT metabolism was only about 7-fold and no general method had been presented how to improve or create new kinase proteins or prodrugs that are suitable for a therapeutic approach. In two other recent reports (Lavie, Nature Struct. Biol. 4 (1997), 601—604; Lavie, Nature Med. 3 (1997), 922—924) it could be shown that the P-loop of the yeast TmpK is involved in limiting the conversion to AZT to AZTTP. The conclusion of crystallographic studies presented in these reports with the yeast TmpK enzyme have, however, been questioned, in particular whether other TmpK enzymes such as the human counterpart have the problem with the P-loop

observed with the yeast enzyme (Kenyon, Nature Struct. Biol. 4 (1997), 595—597). Hence, each enzyme seems to present a different situation, and no general method has been presented for modeling the catalytic domain of a kinase enzyme so as to obtain a high enzymatic activity with nucleoside or nucleotide analogs, such as AZT; nor was it even clear to what extent such method was required or possible.

Thus, the technical problem of the present invention is to provide means and methods for the preparation and activation of nucleoside and nucleotide based drugs.

The solution to this technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the invention relates to a method for the production of a polypeptide having or having enhanced kinase activity for a nucleoside or nucleotide analog, said method comprising substituting, adding or deleting at least one amino acid of a protein having nucleoside or nucleotide kinase activity at a position in the protein where:

(a) the amino acid is at position X_2 and/or X_3 in the consensus sequence $GX_1X_2X_3X_4GK$ of the P-loop;

(b) the amino acid is in the LID region; and/or

(c) the amino acid is at position 105 in the amino acid sequence of human thymidylate kinase or at a corresponding position in a protein having nucleoside or nucleotide kinase activity.

In context with the present invention, the term "nucleoside or nucleotide analog" refers to naturally occurring nucleosides that are modified at the sugar/base while nucleotide analogs may additionally have modifications, e.g., of the α-phosphate. Usually, said nucleoside is adenosine, cytidine, guanosine, thymidine or uridine or based on any of these, such as inosine. Nucleoside and nucleotide analogs useful in accordance with the present invention include 2', 3'-dideoxynucleoside analogs

and other analogs lacking a hydroxyl group equivalent to the 3'-OH of natural nucleoside in terms of their reactivity with reverse transcriptase or other DNA polymerases. Preferably said nucleotide is a nucleoside monophosphate, most preferably said nucleoside monophosphate is thymidylate. A prominent example for a nucleoside (thymidine) analog is 3'-azido-3'-deoxythymidine (AZT) used in the treatment of AIDS.

The term "kinase activity for a nucleoside or nucleotide analog", within the meaning of the present invention refers to the capability of an enzyme to catalyze the phosphorylation of a nucleoside or nucleotide analog to its corresponding mono-, di- or triphosphate. Naturally, such enzymes, also termed kinases herein, transfer a phosphoryl group from a donor molecule (usually ATP) to an acceptor molecule (in the present invention a nucleoside, nucleotide monophosphate, or a nucleotide diphosphate). In the case of thymidylate kinase, the physiological substrates are ATP and dTMP resulting in ADP and dTDP. AZT is very similar to dTMP but has an azido moiety at the 3' position instead of a hydroxyl group, and is thus also activated by the same enzymes that activate thymidine to thymidine triphosphate (dTTP, a substrate for DNA polymerases).

The term "P-loop", as used herein means a motif that has been identified in many ATP- and GTP-binding proteins (Saraste et al., 1990). The binding of the nucleoside or nucleotide to the P-loop is through main-chain nitrogen atoms to the phosphates of the nucleotide, and through a strictly conserved lysine. A comparison of the amino acid sequences of, for example, thymidylate kinases reveal the consensus sequence recited in (a) above. However, other kinases display this motif as well and could thus be used in accordance of the present invention.

The term "LID region" refers to the mobile part of the kinase enzyme that in case of so called class II (type II) TmpKs (bacterial TmpKs such as from *E. coli*) carries one or more catalytically important arginine or lysine residues and undergoes substantial structural rearrangement upon substrate binding. Accordingly, the LID region of class I (type I) TmpKs (eukaryotic enzymes such as human or yeast

TmpK) constitutes the same flexible region without the catalytically important arginine or lysine residues. As will be appreciated by the person skilled in the art from the disclosure herein the P-loop and/or the LID regions of class II (type II) and class I (type I) enzymes can be transferred into each other via, e.g. amino acid substitution(s), additions(s) and/or deletion(s). Thus, according to the method of the present invention, any protein having nucleoside or nucleotide kinase activity can be used and modified so as to obtain a polypeptide which displays said kinase activity for a nucleoside or nucleotide analog. Moreover, if said protein already displays such activity at a basal level, it is possible to improve its kinase activity for nucleoside and nucleotide analogs compared to the wildtype protein. Furthermore, it is possible using the teaching of the present invention to confer to a polypeptide kinase activity for a nucleoside and nucleotide analog via, e.g., protein design using, for example, computer based redesign of proteins. The resultant proteins and polypeptides obtainable by the method of the present invention can then be tested for their kinase activity by using methods known in the art as described, e.g., in the appended examples or in Guettari (1997).

The present invention is based on the following observations. AZT is a prodrug widely used in the treatment of HIV infection. While the concentration of the active form of AZT, AZT triphosphate (AZT-TTP), reaches only the low micromolar range in human cells exposed to AZT, AZT monophosphate (AZT-MP) accumulates to millimolar concentration. Thus, not only are suboptimal concentrations of the active form of the drug produced, but as a further consequence, a highly toxic intermediate (AZT-MP) is produced at high concentration. The reason behind this accumulation of AZT-MP is the poor phosphorylation of AZT-MP to AZT-DP by the enzyme thymidylate kinase (Furman, Proc. Natl. Acad. Sci. USA 83 (1986), 8333-8337).

In accordance with the present invention, surprisingly it has been found that the reasons behind this low activity of the enzyme is to be found in the interplay of two motifs, the P-loop and the LID-region. In experiments performed in accordance with the present invention, the crystal structure of yeast thymidylate kinase (TmpK)

complexed with the bisubstrate inhibitor P1-(5'-adenosyl) P5-(5'-thymidyl) pentaphosphate (TP_5A) was determined at 2.0 Å resolution. In this complex, TmpK adopts a closed conformation with a region (LID) of the protein closing upon the substrate and forming a helix. The interactions of TmpK and TP_5A revealed that arginine 15, which is located in the phosphate binding loop (P-loop) sequence, plays a catalytic role by interacting with an oxygen atom of the transferred phosphoryl group. Unlike other nucleoside monophosphate kinases where basic residues from the LID region participate in stabilizing the transition state, class I (type I) TmpK lack such residues in the LID region. The present inventors attribute this function to Arg15 of the P-loop. TmpK plays an important role in the phosphorylation of the AIDS prodrug AZT. The structures of TmpK with dTMP and with AZT-MP (Lavie et al, Nature Structural Biology 4 (1997), 601-604) implicate the movement of the Arg15 in response to AZT-MP binding as an important factor for the 200 fold reduced catalytic rate with AZT-MP. TmpK from *E. coli* lacks this arginine in its P-loop while having basic residues in the LID region. This suggested that if such a P-loop movement were to occur in the *E. coli* TmpK upon AZT-MP binding, it should not have such a detrimental effect on catalysis. This hypothesis was tested and as expected by the inventors *E. coli* TmpK phosphorylates AZT-MP only 2.5 times slower than dTMP.

Based on the understanding gained from the present work it became possible to mutate the wild type enzyme in order to attain higher activity for AZT-MP. With the advent of gene therapy, it is now conceivable to transfer the gene of such a modified kinase to the HIV-susceptible cells in an infected individual. The subsequent administration of AZT should result in a higher therapeutic index compared with an individual possessing only the wild type kinase.

The next few paragraphs will further illustrate the invention by way of example for TmpKs and describe how kinases catalyze the phosphoryl transfer reaction, and list the possible changes the person skilled in the art can consider in accordance with the present invention, to achieve (higher) kinase activity for nucleoside or nucleotide analogs.

Kinases transfer a phosphoryl group from a donor molecule (usually ATP) to an acceptor molecule (in the present invention a nucleoside, nucleotide monophosphate, or a nucleotide diphosphate). In the case of thymidylate kinase, the physiological substrates are ATP and dTMP resulting in ADP and dTDP. AZT is very similar to dTMP but has an azido moiety at the 3' position instead of an hydroxyl group, and is thus also activated by the same enzymes that activate thymidine to thymidine triphosphate (dTTP, a substrate for DNA polymerases). Enzymes catalyze chemical reactions by preferentially stabilizing the transition state of a reaction over the corresponding ground state. Since it is highly probable that in phosphoryl transfer reactions an additional negative charge is formed at the transition state in comparison to the total charge of the ground state, stabilizing such a charge lowers the energy barrier for the reaction. Enzymes can stabilize negative charges either by using positively charged amino acid residues (Arg, Lys, or His) or by binding metals, or both. Kinases such as thymidylate kinases utilize both of these possibilities, but since the interaction with the metal (a bound magnesium ion) is probably similar in the ground and transition states, the energy barrier between ground and transition state is not appreciably decreased (this does not mean that the metal is not important for catalysis; the absence of metal abolishes all activity). It is the positive side chain of arginine residue(s) that change their degree of interaction between the ground and transition states and thus achieve preferential stabilization of the transition state over the ground state (i.e. a lowering of the energy barrier for the phosphoryl transfer).

The structure of thymidylate kinase complexed with the bisubstrate analog TP_5A that has been solved in accordance with the present invention allows to observe which basic residues interact with the nucleotides. To date a unique feature of class I (type I) TmpKs is that an arginine from the P-loop interacts directly with the transferred phosphoryl group. The present inventors postulate that this interaction is responsible for the low activity of TmpK with AZT-MP since the azido group of AZT pushes the P-loop from its position in comparison to its position when dTMP is bound.

An important observation made in accordance with the present invention is that the *E. coli* TmpK (a class II (type II) enzyme) phosphorylates AZT-MP very rapidly. The *E. coli* TmpK has no arginine in its P-loop (a glycine instead) but rather a number of basic resides in a structural motif that is called LID region. This is an example of the scaffold nature of enzymes: Class I (type I) TmpKs have a basic residue in the P-loop motif but none in the LID, whereas class II (type II) TmpKs have the reverse; i.e. none in the P-loop and a few in the LID. In other words, class I (type I) kinases use the P-loop arginine to stabilize the transition state whereas class II (type II) use arginine(s) from the LID for the same purpose.

The above result has immediate implications. Instead of modifying the human or other class I (type I) kinase to better phosphorylate AZT-MP, the *E. coli* TmpK could be used for this purpose (and possibly any other class II (type II) TmpK). The kinetic results described in the appended examples demonstrate that the *E. coli* TmpK is at least 300-fold faster in phosphorylating AZT-MP than the human TmpK. However, because of immunological and other reasons it might be advantageous to use a modified human TmpK over the *E. coli* TmpK in the proposed gene therapy procedure. The modification of the human (or any other class I (type I)) TmpK would be according to the method invention as described above. Advantageously the method of the invention comprises at least one of the following steps:

- Mutation of the LID region of class I (type I) thymidylate kinases to contain one or more basic residues and in doing so mimic the LID motif of class II (type II) TmpKs. This mutation of the LID could be with or without a concomitant mutation in the P-loop. A concomitant substitution in the P-loop (arginine to a smaller amino acid) might be advantageous in order to avoid a steric clash between the P-loop basic residue and the newly introduced LID basic residue(s). The results obtained in accordance with the present invention demonstrate that such an approach is appropriate but they do not rule out that, for example, a lysine in the P-loop might be tolerated.

- Substitution of the entire LID of a class I (type I) TmpK with that from a class II (type II) TmpK, with or without a concomitant mutation in the P-loop; see, e.g., Example 8.
- Selection of appropriate mutant enzymes by measurements of the catalytic efficiency with a coupled colorometric assay that links the phosphorylation of, e.g., the nucleoside monophosphate to diphosphate to a change of the optical density. This can be extended to a high throughput screening (HTS) setup with a microtiter plate reader.

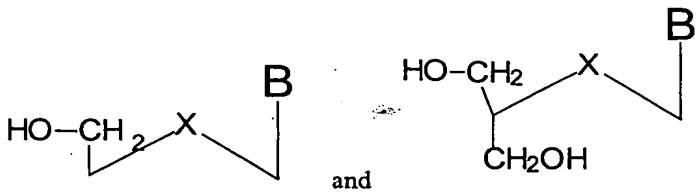
The method of the invention can be performed using conventional techniques known in the art, for example, by using amino acid deletion(s), insertion(s), substitution(s), addition(s), and/or recombination(s) and/or any other modification(s) known in the art either alone or in combination. Methods for introducing such modifications in the DNA sequence underlying the amino acid sequence of a protein having kinase activity are well known to the person skilled in the art; see, e.g., Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y.

In a preferred embodiment of the method of the invention said protein to be modified with the method described above is derived from a eukaryotic or prokaryotic organism, preferably said organism is human or yeast. As described above, the present inventors have solved the 3-dimensional structure of TmpK complexed with the bisubstrate inhibitor P1-(5'-Adenosyl)P5-(5'-Thymidyl) pentaphosphate (TP_5A). This structure, taken together with those of TmpK complexed with dTMP and with AZT-MP (Lavie et al., 1997b) allowed the inventors to identify the residues important for catalysis, and to propose an explanation for the slow phosphorylation of AZT-MP. In addition, based on that knowledge, it permits the prediction that the rate of AZT-MP phosphorylation by the *E. coli* TmpK would be similar to the phosphorylation rate of the physiological substrate dTMP. In fact, as shown in accordance with the present invention *E. coli* TmpK discriminates poorly against AZT-MP, supporting the inventors finding that Arg15

of yeast and human thymidylate kinases is responsible for discrimination against AZT-MP as a substrate (the human thymidylate kinase has 44% amino acid identity and >63% similarity with the yeast enzyme with all catalytically important residues conserved between the two species; see Figure 4). Furthermore, as demonstrated in Example 8 herein below the yeast TmpK could, by applying the method of the invention, converted into an enzyme displaying high kinase activity for the nucleoside analog AZT. Similar results are to be expected when using the human or any other corresponding eukaryotic enzyme.

In a preferred embodiment of the method of the invention said protein comprises the amino acid sequence of any one of SEQ ID NOS: 1 to 13 or a fragment thereof. The amino acid sequenes depicted in SEQ ID NOS: 1 to 13 belong to various TmpKs of eukaryotic and prokaryotic origin and comprise, for example, the amino acid sequence of *E. coli* TmpK. With the method of the present invention and the teaching provided herein, it is, e.g., possible to improve the kinase activity of the human or yeast enzyme for nucleoside or nucleotide analogs, in particular for AZT. Furthermore, the teaching of the present invention now enables rationale shuffling or domains from, e.g., *E. coli* TmpK to the corresponding enzyme from human, mouse or yeast. On the other hand it is also possible to modify the amino acid sequence of the *E. coli* enzyme to more closely resemble that of the corresponding eukaryotic, preferably human enzyme while the P-loop and the LID region remain substanstially unaffected and therefore the resultant polypeptide retains its kinase activity for nucleoside and nucleotide analogs. For the rational design of polypeptides produced according to the method of the invention computer programs may be used such as BRASMOL that are obtainable from the Internet. Furthermore, folding simulations and computer redesign of structural motifs can be performed using other appropriate computer programs (Olszewski, Proteins 25 (1996), 286-299; Hoffman, Comput. Appl. Biosci. 11 (1995), 675-679). Computers can be used for the conformational and energetic analysis of detailed protein models (Monge, J. Mol. Biol. 247 (1995), 995-1012; Renouf, Adv. Exp. Med. Biol. 376 (1995), 37-45).

In another preferred embodiment the method of of the present invention said amino acid which is substituted or added in (a) is glycine or lysine. In case that the amino acid is changed in the P-loop of class I (type I) enzymes the amino acid that is changed is preferably a small amino acid such as glycine. However, alanine, less preferably valine, threonine, serine or glutamic acid may work as well although probably less effective. As described above, the LID region of class I (type I) thymidylate kinases may be substituted in accordance with the method of the invention to contain at least one basic residue. Thus, in a preferred embodiment of the method of the invention said amino acid which is substituted or added in (b) is a basic amino acid, preferably arginine. Usually, said LID has the consensus sequence R/KXXXXXERYEXXXXQ. This consensus sequence or substantially identical sequences can be determined by comparison of the amino acid sequences of known nucleoside and nucleotide kinases, for example those shown in SEQ ID NOS: 1 to 13. Preferably, said polypeptide obtainable by the method of the invention exhibits kinase activity for a nucleoside or nucleotide analog which is higher than that of the corresponding wild type enzyme, preferably higher than that of an eukaryotic, most preferably higher than that of the human enzyme. Preferably, said kinase activity for a nucleoside or nucleotide analog is 5-fold, more preferably 10-fold, still more preferably 30-fold and most preferably 300-fold improved compared to the corresponding wild-type enzyme. Furthermore, it is preferred within the method of the invention that said nucleoside analog is AZT, d4T or has the following structure:



wherein B is any nucleobase or analog thereof, and X is O, CH_2, NH or S.

As discussed above, the amino acid substitution(s) or other modifications performed in the amino acid sequence of a given kinase may result in the P-loop and/or in the LID region of a bacterial nucleoside or nucleotide kinase, preferably those of the TmpK of E.coli; see also Example 8.

As mentioned before, for the long term aim of gene therapeutic potentiation of AZT effectiveness, there could be significant advantages from use of a modified human enzyme. Therefore, the mutational studies were extended to the human enzyme; see Example 9. Contrary to all expectations arising from the studies on the yeast enzyme, replacement of Arg-16 (equivalent to Arg-15 in the yeast enzyme) did not lead to a loss in catalytic activity, but to a slight gain. It is possible that this is related to the fact that the k_{cat} for the human enzyme is much slower than for the yeast enzyme (0.67 cf. 35 s^{-1}), which is already a very slow kinase when compared with other nucleoside monophosphate kinases. Without intending to be bound by theory it is believed that since the catalytic machinery in both enzymes appears to be identical, as shown by sequence comparison and 3-D structural determination, it is possible that the chemical step (i.e. phosphate transfer) is not rate limiting, but rather product release, so that slowing down a relatively rapid chemical step might not have any influence on the overall rate.

In keeping with expectations arising from the yeast TMPK data was the fact that introduction of the E. coli lid region without removing Arg-16 led to a drop in catalytic activity. However, and most dramatically, a combination of introduction of the E. coli lid with replacement of Arg-16 by glycine not only restored more than full wild type catalytic activity, but resulted in a protein which is more efficient with AZT-MP than with TMP. The increase in activity with AZT-MP is approximately 300 fold. This mutant thus has properties which are highly attractive for improving the potency of AZT, since the efficiency of AZTMP phosphorylation is improved dramatically with only a minor increase in TMP phosphorylation activity. The latter is an important aspect, since AZTTP must compete with TTP for HIV-reverse transcriptase catalysed addition to the end of a growing (HIV) DNA chain.

The mutants of TMPK described so far showing altered specificity for AZTMP and TMP were produced according to rational considerations based on comparative structure-function studies of TMPK from three different sources. In this respect, a further striking result is that obtained by replacing Phe-105 in the human enzyme by tyrosine. The rational for this was that in the yeast enzyme, the corresponding residue (at position 102) is a tyrosine, and its hydroxyl group interacts with the carboxylate side chain of Asp-14, which (like Asp-15 in the human enzyme) appears to be an essential residue for the catalytic mechanism. The Phe-105 Tyr mutant of the human enzyme shows reduced activity with TMP, but, unexpectedly, significantly increased activity with AZTMP, so that the latter is now a better substrate than the former; see Example 9. Thus, and again without intending to be bound by theory it is believed that substitution of amino acid(s) that interact with the LID or P-loop, and in particular, if present, with the carboxylate side chain of Asp leads to a conformational change of the catalytic center of the kinase enzyme and fulfils the requirement of greatly increased AZTMP phosphorylating activity without increasing (actually decreasing) TMP phosphorylation. An amino acid position in a nucleotide or nucleoside kinase corresponding to amino acid residue 105 in the human thymidylate kinase can be determined by comparison and alignment of the amino acid sequences of known nucleoside and nucleotide kinases, for example, those shown in SEQ ID NOS: 1 to 13 using, e.g., the program GCG. The specificity ratio (AZTMP:TMP) is actually identical with that of the mutant containing Gly-16 and the E. coli lid, although the overall activity is a factor of ca. 10 lower.

The results obtained in accordance with the present invention show how rational considerations have led to generation of TMPK mutants capable of phosphorylating the nucleoside pro-drug AZT with high efficiency. This finding gives rise to several applications in the medical and diagnostic field which will be explained in more detail below.

In another embodiment the present invention relates to a polynucleotide encoding the polypeptide obtainable by the method of the invention. Said polynucleotide

may be, e.g., DNA, cDNA, genomic DNA, RNA or synthetically produced DNA or RNA or a recombinantly produced chimeric nucleic acid molecule comprising any of those polynucleotides either alone or in combination. Preferably said polynucleotide is part of a vector. Such vectors may comprise further genes such as marker genes which allow for the selection of said vector in a suitable host cell and under suitable conditions.

In a further preferred embodiment the polynucleotide of the invention is operatively linked to expression control sequences allowing expression in prokaryotic or eukaryotic cells.

Expression of said polynucleotide comprises transcription of the polynucleotide into a translatable mRNA. Regulatory elements ensuring expression in eukaryotic cells, preferably mammalian cells, are well known to those skilled in the art. They usually comprise regulatory sequences ensuring initiation of transcription and optionally poly-A signals ensuring termination of transcription and stabilization of the transcript. Additional regulatory elements may include transcriptional as well as translational enhancers. Possible regulatory elements permitting expression in prokaryotic host cells comprise, e.g., the *lac*, *trp* or *tac* promoter in *E. coli*, and examples for regulatory elements permitting expression in eukaryotic host cells are the *AOX1* or *GAL1* promoter in yeast or the CMV-, SV40- , RSV-promoter (Rous sarcoma virus), CMV-enhancer, SV40-enhancer or a globin intron in mammalian and other animal cells. Beside elements which are responsible for the initiation of transcription such regulatory elements may also comprise transcription termination signals, such as the SV40-poly-A site or the tk-poly-A site, downstream of the polynucleotide. In this context, suitable expression vectors are known in the art such as Okayama-Berg cDNA expression vector pcDV1 (Pharmacia), pCDM8, pRc/CMV, pcDNA1, pcDNA3 (In-vitrogene), pSPORT1 (GIBCO BRL).

The polynucleotide of the invention can be used alone or as part of a vector to express the polypeptide having kinase activity for nucleoside or nucleotide analogs in cells, for, e.g., gene therapy or diagnostics of diseases related to viral infections

and cancer. The polynucleotide or vector containing the DNA sequence encoding a kinase for nucleoside or nucleotide analogs is introduced into the cells which in turn produce the protein of interest. Gene therapy, which is based on introducing therapeutic genes into cells by *ex-vivo* or *in-vivo* techniques is one of the most important applications of gene transfer. Suitable vectors and methods for *in-vitro* or *in-vivo* gene therapy are described in the literature and are known to the person skilled in the art; see, e.g., Giordano, Nature Medicine 2 (1996), 534-539; Schaper, Circ. Res. 79 (1996), 911-919; Anderson, Science 256 (1992), 808-813; Isner, Lancet 348 (1996), 370-374; Muhlhauser, Circ. Res. 77 (1995), 1077-1086; Wang, Nature Medicine 2 (1996), 714-716; WO94/29469; WO 97/00957 or Schaper, Current Opinion in Biotechnology 7 (1996), 635-640, and references cited therein. The polynucleotides and vectors of the invention may be designed for direct introduction or for introduction via liposomes, or viral vectors (e.g. adenoviral, retroviral) into the cell. Preferably, said cell is a germ line cell, embryonic cell, or egg cell or derived therefrom, most preferably said cell is a stem cell.

Furthermore, the present invention relates to vectors, particularly plasmids, cosmids, viruses and bacteriophages used conventionally in genetic engineering that comprise a polynucleotide of the invention. Preferably, said vector is an expression vector and/or a gene transfer or targeting vector. Expression vectors derived from viruses such as retroviruses, vaccinia virus, adeno-associated virus, herpes viruses, or bovine papilloma virus, may be used for delivery of the polynucleotides or vector of the invention into targeted cell population. Methods which are well known to those skilled in the art can be used to construct recombinant viral vectors; see, for example, the techniques described in Sambrook, Molecular Cloning A Laboratory Manual, Cold Spring Harbor Laboratory (1989) N.Y. and Ausubel, Current Protocols in Molecular Biology, Green Publishing Associates and Wiley Interscience, N.Y. (1989). Alternatively, the polynucleotides and vectors of the invention can be reconstituted into liposomes for delivery to target cells.

The present invention furthermore relates to host cells transformed with a polynucleotide or vector of the invention. Said host cell may be a prokaryotic or eukaryotic cell. The polynucleotide or vector of the invention which is present in the host cell may either be integrated into the genome of the host cell or it may be maintained extrachromosomally. In this respect, it is also to be understood that the recombinant DNA molecule of the invention can be used for "gene targeting" and/or "gene replacement", for restoring a mutant gene or for creating a mutant gene via homologous recombination; see for example Mouellic, Proc. Natl. Acad. Sci. USA, 87 (1990), 4712-4716; Joyner, Gene Targeting, A Practical Approach, Oxford University Press.

The host cell can be any prokaryotic or eukaryotic cell, such as a bacterial, insect, fungal, plant, animal or human cell. Preferred fungal cells are, for example, those of the genus Saccharomyces, in particular those of the species S. cerevisiae. The term "prokaryotic" is meant to include all bacteria which can be transformed or transfected with a polynucleotide for the expression of a polypeptide of the invention. Prokaryotic hosts may include gram negative as well as gram positive bacteria such as, for example, *E. coli*, *S. typhimurium*, *Serratia marcescens* and *Bacillus subtilis*. A polynucleotide coding for a polypeptide of the invention can be used to transform or transfect the host using any of the techniques commonly known to those of ordinary skill in the art. Especially preferred is the use of a plasmid or a virus containing the coding sequence of the polypeptide having kinase activity for a nucleoside or nucleotide analog for purposes of prokaryotic transformation or transfection, respectively. Methods for preparing fused, operably linked genes and expressing them in bacteria are well-known in the art (Maniatis, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989). The genetic constructs and methods described therein can be utilized for expression of the polypeptide of the invention in prokaryotic hosts. In general, expression vectors containing promoter sequences which facilitate the efficient transcription of the inserted polynucleotide are used in connection with the host. The expression vector typically contains an origin of replication, a promoter, and a terminator, as well as specific genes which are

capable of providing phenotypic selection of the transformed cells. The transformed prokaryotic hosts can be grown in fermentors and cultured according to techniques known in the art to achieve optimal cell growth. The polypeptides of the invention can then be isolated from the grown medium, cellular lysates, or cellular membrane fractions. The isolation and purification of the microbially expressed polypeptides of the invention may be by any conventional means such as, for example, preparative chromatographic separations and immunological separations such as those involving the use of monoclonal or polyclonal antibodies.

Thus, in a further embodiment the invention relates to a method for the production of a polypeptide having kinase activity for a nucleoside or nucleotide analog comprising culturing a host cell as defined above under conditions allowing the expression of the polypeptide and recovering the produced polypeptide from the culture.

In another embodiment the present invention relates to a method for producing cells capable of expressing a polypeptide having nucleoside or nucleotide kinase activity for nucleoside or nucleotide analogs comprising genetically engineering cells with the polynucleotide or with the vector of the invention. The cells obtainable by the method of the invention can be used, for example, to test the capacity of the polypeptides of the invention to improve the metabolism of nucleoside or nucleotide analogs and, if mammalian cells are used, to test their impact on anti-viral activity of nucleoside and nucleotide analogs. Furthermore, the cells can be used to study known and unkown nucleoside and nucleotide analogs for their ability to be converted to the corresponding mono-, di- or triphosphates. The cells obtainable by the above-described method may also be used for the screening methods referred to herein below.

Furthermore, the invention relates to a polypeptide having kinase activity for a nucleoside or nucleotide analog encoded by a polynucleotide according to the

invention or obtainable by the above-described methods or from cells produced by the method described above.

In this context it is also understood that the polypeptides according to the invention may be further modified by conventional methods known in the art. By providing the polypeptides according to the present invention it is also possible to determine the portions relevant for their biological activity, namely their kinase activity. This may allow the construction of chimeric proteins comprising an amino acid sequence derived from a polypeptide of the invention which is crucial for kinase activity and other functional amino acid sequences e.g. nuclear localization signals, transactivating domains, DNA-binding domains, hormone-binding domains, protein tags (GST, GFP, h-myc peptide, Flag, HA peptide) which may be derived from the same or from heterologous proteins.

The present invention furthermore relates to antibodies specifically recognizing a polypeptide according to the invention which has kinase activity for a nucleoside or nucleotide analog. Advantageously, the antibody specifically recognizes a polypeptide according to the invention which has kinase activity for a nucleoside or nucleotide analog but does not recognize a polypeptide which is a wild type starting protein of such a polypeptide and which has no or less kinase activity for nucleoside or nucleotide analogs than the polypeptides of the invention.
Antibodies against the polypeptide of the invention can be prepared by well known methods using a purified polypeptide according to the invention or a synthetic fragment derived therefrom as an antigen. Monoclonal antibodies can be prepared, for example, by the techniques as originally described in Köhler and Milstein, Nature 256 (1975), 495, and Galfré, Meth. Enzymol. 73 (1981), 3, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals. The antibodies can be monoclonal antibodies, polyclonal antibodies or synthetic antibodies as well as fragments of antibodies, such as Fab, Fv or scFv fragments etc. Furthermore, antibodies or fragments thereof to the aforementioned polypeptides can be obtained by using methods which are

described, e.g., in Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988. These antibodies can be used, for example, for the immunoprecipitation and immunolocalization of the polypeptides of the invention as well as for the monitoring of the presence of such polypeptides, for example, in recombinant organisms, and for the identification of compounds interacting with the proteins according to the invention. For example, surface plasmon resonance as employed in the BIAcore system can be used to increase the efficiency of phage antibodies which bind to an epitope of the polypeptide of the invention (Schier, Human Antibodies Hybridomas 7 (1996), 97-105; Malmborg, J. Immunol. Methods 183 (1995), 7-13).

Moreover, the present invention relates to a composition, preferably a pharmaceutical composition comprising

(a) a prokaryotic protein having nucleoside or nucleotide kinase activity for a nucleoside or nucleotide analog or a polynucleotide encoding and capable of expressing said protein *in vivo* or a vector containing said polynucleotide; or

(b) the above-described polypeptide, polynucleotide or vector of the invention;

(c) optionally a nucleoside or nucleotide analog; and

(d) optionally a pharmaceutically acceptable carrier.

As described above and as shown in the appended examples prokaryotic kinases such as *E. coli* TmpK belong to class II (type II) enzymes which display superior phosphorylation properties for nucleoside and nucleotide analogs, in particular for AZT. Thus, it is not necessary to further modify such prokaryotic proteins according to the method of the invention but it is possible to employ them in a pharmaceuitcal composition of the invention unmodified or substantially unmodified.

Examples of suitable pharmaceutical carriers are well known in the art and include phosphate buffered saline solutions, water, emulsions, such as oil/water emulsions, various types of wetting agents, sterile solutions etc. Compositions comprising such carriers can be formulated by well known conventional methods.

These pharmaceutical compositions can be administered to the subject at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g., by intravenous, intraperitoneal, subcutaneous, intramuscular, topical or intradermal administration. The dosage regimen will be determined by the attending physician and other clinical factors. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Generally, the regimen as a regular administration of the pharmaceutical composition should be in the range of 1 µg to 10 mg units per day. If the regimen is a continuous infusion, it should also be in the range of 1 µg to 10 mg units per kilogram of body weight per minute, respectively. Progress can be monitored by periodic assessment. Dosages will vary but a preferred dosage for intravenous administration of DNA is from approximately 10^6 to 10^{12} copies of the DNA molecule. The compositions of the invention may be administered locally or systemically. Administration will generally be parenterally, e.g., intravenously; DNA may also be administered directly to the target site, e.g., by biolistic delivery to an internal or external target site or by catheter to a site in an artery.

It is envisaged by the present invention that the various polynucleotides and vectors of the invention are administered either alone or in any combination using standard vectors and/or gene delivery systems, and optionally together with an appropriate compound, for example a nucleoside or nucleotide analog, and/or together with a pharmaceutically acceptable carrier or excipient. Subsequent to administration, said polynucleotides or vectors may be stably integrated into the genome of the subject. On the other hand, viral vectors may be used which are specific for certain cells or tissues, preferably for CD4 cells and persist in said cells. Suitable pharmaceutical carriers and excipients are well known in the art. The pharmaceutical compositions prepared according to the invention can be used for the prevention or treatment or delaying of different kinds of diseases, which are related to viral infection or cancer.

Furthermore, it is possible to use a pharmaceutical composition of the invention which comprises polynucleotide or vector of the invention in gene therapy. Suitable gene delivery systems may include liposomes, receptor-mediated delivery systems, naked DNA, and viral vectors such as herpes viruses, retroviruses, adenoviruses, and adeno-associated viruses, among others. Delivery of nucleic acids to a specific site in the body for gene therapy may also be accomplished using a biolistic delivery system, such as that described by Williams (Proc. Natl. Acad. Sci. USA 88 (1991), 2726-2729).

Standard methods for transfecting cells with recombinant DNA are well known to those skilled in the art of molecular biology, see, e.g., WO 94/29469. Gene therapy may be carried out by directly administering the recombinant DNA molecule or vector of the invention to a patient or by transfecting cells such as CD4 with the polynucleotide or vector of the invention ex vivo and infusing the transfected cells into the patient. Furthermore, research pertaining to gene transfer into cells of the germ line is one of the fastest growing fields in reproductive biology. Gene therapy, which is based on introducing therapeutic genes into cells by ex-vivo or in-vivo techniques is one of the most important applications of gene transfer. Suitable vectors and methods for in-vitro or in-vivo gene therapy are described in the literature and are known to the person skilled in the art; see, e.g., WO 94/29469, WO 97/00957 or Schaper (Current Opinion in Biotechnology 7 (1996), 635-640) and references cited above. The polynucleotides and vectors comprised in the pharmaceutical composition of the invention may be designed for direct introduction or for introduction via liposomes, or viral vectors (e.g. adenoviral, retroviral) containing said recombinant DNA molecule into the cell. Preferably, said cell is a germ line cell, embryonic cell, stem cell or egg cell or derived therefrom. The pharmaceutical compositions according to the invention can be used for the treatment of diseases hitherto unknown as being related to viral infection or cancer. An embryonic cell can be for example an embryonic stem cell as described in, e.g., Nagy, Proc. Natl. Acad. Sci. USA 90 (1993) 8424-8428.

It is to be understood that the introduced polynucleotides and vectors of the invention express the polypeptide or protein having kinase activity for nucleoside or nucleotide analogs after introduction into said cell and preferably remain in this status during the lifetime of said cell. For example, cell lines which stably express the polynucleotide under the control of appropriate regulatory sequences may be engineered according to methods well known to those skilled in the art. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with the polynucleotide or vector of the invention and a selectable marker, either on the same or separate vectors. Following the introduction of foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows for the selection of cells having stably integrated the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. Such engineered cell lines are particularly useful in screening methods described below.

A number of selection systems may be used, including but not limited to the herpes simplex virus thymidine kinase (Wigler, Cell 11(1977), 223), hypoxanthine-guanine phosphoribosyltransferase (Szybalska, Proc. Natl. Acad. Sci. USA 48 (1962), 2026), and adenine phosphoribosyltransferase (Lowy, Cell 22 (1980), 817) in tk^-, $hgprt^-$ or $aprt^-$ cells, respectively. Also, antimetabolite resistance can be used as the basis of selection for dhfr, which confers resistance to methotrexate (Wigler, Proc. Natl. Acad. Sci. USA 77 (1980), 3567; O'Hare, Proc. Natl. Acad. Sci. USA 78 (1981), 1527), gpt, which confers resistance to mycophenolic acid (Mulligan, Proc. Natl. Acad. Sci. USA 78 (1981), 2072); neo, which confers resistance to the aminoglycoside G-418 (Colberre-Garapin, J. Mol. Biol. 150 (1981), 1); hygro, which confers resistance to hygromycin (Santerre, Gene 30 (1984), 147); or puromycin (pat, puromycin N-acetyl transferase). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan; hisD, which allows cells to utilize histinol in place of histidine (Hartman, Proc. Natl. Acad. Sci. USA 85 (1988), 8047); and ODC (ornithine decarboxylase)

which confers resistance to the ornithine decarboxylase inhibitor, 2-(difluoromethyl)-DL-ornithine, DFMO (McConlogue, 1987, In: Current Communications in Molecular Biology, Cold Spring Harbor Laboratory ed.). Cells to be used for *ex vivo* gene therapy are well known to those skilled in the art. For example, mature cells of the immune system present in blood or preferably the corresponding stem cells.

In a preferred embodiment of the invention said protein in the pharmaceutical composition is a bacterial nucleoside or nucleotide kinase, preferably a bacterial TmpK. As described in the appended examples the TmpK of *E. coli* or its structure with respect to the P-loop and the LID region appear to be particularly suited for the activation of nucleoside and nucleotide analogs, e.g., AZT. Thus, in a preferred embodiment of the pharmaceutical composition of the invention said protein to be employed in accordance with the present invention has at least the P-loop and/or the LID region of E.coli TmpK, for example, said TmpK comprises the amino acid sequence of *E. coli* TmpK, e.g., that shown in SEQ ID NO: 4 or a biologically active fragment thereof.

The present invention also relates to compositions comprising at least one of the aforementioned polynucleotides, vectors, polypeptides, proteins or antibodies, and in the case of kits or diagnostic compositions, optionally suitable means for detection. Said compositions may further contain compounds such as nucleoside or nucleotide analogs, further plasmids, antibiotics and the like for screening transgenic cells useful for the genetic engineering of non-human animals, preferably mammals and most preferably mouse. The diagnostic compositions of the invention may be used for methods of detecting and isolating nucleoside or nucleotide analogs which are functionally equivalent to, e.g., AZT or d4T.

The present invention also relates to a method for the production of a transgenic non-human animal, preferably transgenic mouse, comprising introduction of a polynucleotide or vector of the invention into a germ cell, an embryonic cell, stem

cell or an egg or a cell derived therefrom. The non-human animal to be used in the method of the invention may be a non-transgenic healthy animal, or may have a viral disease or cancer, preferably a disease caused by infection of a retrovirus such as HIV, HTLV or related viruses. Production of transgenic embryos and screening of those can be performed, e.g., as described by A. L. Joyner Ed., Gene Targeting, A Practical Approach (1993), Oxford University Press. The DNA of the embryonal membranes of embryos can be analyzed using Southern blots with an appropriate probe.

The invention also relates to transgenic non-human animals such as transgenic non-human mouse, rats, hamsters, dogs, monkeys, rabbits or pigs comprising a polynucleotide or vector of the invention or obtained by the method described above, preferably wherein said polynucleotide or vector is stably integrated into the genome of said non-human animal, preferably such that the presence of said polynucleotide or vector leads to the expression of the polypeptide of the invention.

With the polypeptides, polynucleotides and vectors of the invention, it is now possible to study *in vivo* and *in vitro* the efficiency of nucleoside and nucleotide analogs. Furthermore, since the polypeptides of the invention provide for optimal or at least improved activation of said nucleoside or nucleotide analog, it is now possible to determine further analogs which may be effective for the treatment of viral diseases or cancer, for example specific tumors or AIDS.

The present invention further relates to a method for identifying an inhibitor of a nucleoside or nucleotide kinase comprising the steps of:

(a) contacting the polypeptide of the invention or a cell expressing said polypeptide in the presence of components capable of providing a detectable signal in response to kinase activity, with a compound to be screened under conditions that permit binding of said compound to the nucleoside or nucleotide kinase, and

(b) detecting presence or absence of a signal generated from the kinase activity of the polypeptide, wherein the absence or decrease of the signal is indicative for an inhibitor of a nucleoside or nucleotide kinase.

Furthermore, the invention relates to a method for identifying a nucleoside or nucleotide based prodrug comprising the steps of

(a) contacting the polypeptide of the invention or a cell expressing said polypeptide in the presence of components capable of providing a detectable signal in response to kinase activity, with a nucleoside or nucleotide analog compound to be screened under conditions that permit kinase activity of said polypeptide, and

(b) detecting presence or absence of a signal generated from the kinase activity of the polypeptide, wherein the presence of a signal is indicative for a putative prodrug.

The term " compound" in a method of the invention includes a single substance or a plurality of substances which may or may not be identical.

Said compound(s) may be comprised in, for example, samples, e.g., cell extracts from, e.g., plants, animals or microorganisms. Furthermore, said compounds may be known in the art but hitherto not known to be capable of inhibiting a nucleoside or nucleotide kinase or not known to be useful as a prodrug, respectively. The plurality of compounds may be, e.g., added to the culture medium or injected into a cell or non-human animal of the invention.

If a sample containing (a) compound(s) is identified in the method of the invention, then it is either possible to isolate the compound from the original sample identified as containing the compound, in question or one can further subdivide the original sample, for example, if it consists of a plurality of different compounds, so as to reduce the number of different substances per sample and repeat the method with the subdivisions of the original sample. It can then be determined whether said sample or compound displays the desired properties, for example, by the methods described herein, the appended examples or in the literature, e.g., Guettari (1997). Depending on the complexity of the samples, the steps described above can be

performed several times, preferably until the sample identified according to the method of the invention only comprises a limited number of or only one substance(s). Preferably said sample comprises substances of similar chemical and/or physical properties, and most preferably said substances are identical. The methods of the present invention can be easily performed and designed by the person skilled in the art, for example in accordance with other cell based assays described in the prior art or by using and modifying the methods as described in the appended examples. Furthermore, the person skilled in the art will readily recognize which further compounds and/or enzymes may be used in order to perform the methods of the invention, for example, enzymes, if necessary, that convert a certain compound into the precursor which in turn represents a substrate for the kinase of the invention. Such adaptation of the method of the invention is well within the skill of the person skilled in the art and can be performed without undue experimentation.

Compounds which can be used in accordance with the present invention include peptides, proteins, nucleic acids, antibodies, small organic compounds, ligands, peptidomimetics, PNAs and the like. Said compounds can also be functional derivatives or analogues of known nucleoside or nucleotide analogs. Methods for the preparation of chemical derivatives and analogues are well known to those skilled in the art and are described in, for example, Beilstein, Handbook of Organic Chemistry, Springer edition New York Inc., 175 Fifth Avenue, New York, N.Y. 10010 U.S.A. and Organic Synthesis, Wiley, New York, USA. Furthermore, said derivatives and analogues can be tested for their effects according to methods known in the art or as described, for example, in the appended examples. Furthermore, peptide mimetics and/or computer aided design of appropriate nucleoside and nucleotide derivatives and analogues can be used, for example, according to the methods described below. Nucleoside and nucleotide analogs comprise molecules having as the basis structure a ribo-, deoxyribo- or dideoxyribonucleoside.

Appropriate computer programs can be used for the identification of interactive sites of a putative inhibitor and the polypeptide of the invention by computer assistant searches for complementary structural motifs (Fassina, Immunomethods 5 (1994), 114-120). Further appropriate computer systems for the computer aided design of protein and peptides are described in the prior art, for example, in Berry, Biochem. Soc. Trans. 22 (1994), 1033-1036; Wodak, Ann. N. Y. Acad. Sci. 501 (1987), 1-13; Pabo, Biochemistry 25 (1986), 5987-5991. The results obtained from the above-described computer analysis can be used in combination with the method of the invention for, e.g., optimizing known nucleoside or nucleotide analogs. Appropriate peptide mimetics and other inhibitors can also be identified by the synthesis of peptide mimetic combinatorial libraries through successive chemical modification and testing the resulting compounds, e.g., according to the methods described herein and in the appended examples. Methods for the generation and use of peptide mimetic combinatorial libraries are described in the prior art, for example in Ostresh, Methods in Enzymology 267 (1996), 220-234 and Dorner, Bioorg. Med. Chem. 4 (1996), 709-715. Furthermore, the three-dimensional and/or crystallographic structure of inhibitors of the polypeptide of the invention can be used for the design of peptide mimetic inhibitors or nucleoside and nucleotide analogs, e.g., in combination with the polypeptide of the invention (Rose, Biochemistry 35 (1996), 12933-12944; Rutenber, Bioorg. Med. Chem. 4 (1996), 1545-1558).

The compounds identified according to the method of the invention, in particular nucleoside and nucleotide analogs are expected to be very beneficial since the prodrugs that have been used so far are only of limited use due to their inefficient metabolism in the subject.

In summary, the present invention provides methods for identifying compounds which inhibit kinase activity as well as compounds that can be used as nucleoside or nucleotide based prodrugs even in the absence of a polypeptide or protein having improved kinase activity for nucleoside and nucleotide analogs.

Compounds found to downregulate the activity of kinases may be used in the treatment of cancer and related diseases. In addition, it may also be possible to specifically inhibit viral kinases, thereby preventing viral infection or viral spread. The compounds identified or obtained according to the method of the present invention are thus expected to be very useful in diagnostic and in particular for therapeutic applications. Hence, in a further embodiment the invention relates to a method for the production of a pharmaceutical composition comprising the steps of

(a) contacting the polypeptide of the invention or a cell expressing said polypeptide in the presence of components capable of providing a detectable signal in response to kinase activity, with a compound to be screened under conditions that permit binding of said compound to the nucleoside or nucleotide kinase, and

(b) detecting presence or absence of a signal generated from the kinase activity of the polypeptide, wherein the absence or decrease of the signal is indicative for an inhibitor of a nucleoside or nucleotide kinase, or

(a') contacting the polypeptide of the invention or a cell expressing said polypeptide in the presence of components capable of providing a detectable signal in response to kinase activity, with a nucleoside or nucleotide analog compound to be screened under conditions that permit kinase activity of said polypeptide, and

(b') detecting presence or absence of a signal generated from the kinase activity of the polypeptide, wherein the presence of a signal is indicative for a putative prodrug; and

(c) formulating the inhibitor identified in step (b) or the nucleoside or nucleotide analog identified in step (b') in a pharmaceutically acceptable form.

The therapeutically useful compounds identified according to the method of the invention may be administered to a patient by any appropriate method for the particular compound, e.g., orally, intravenously, parenterally, transdermally, transmucosally, or by surgery or implantation (e.g., with the compound being in the form of a solid or semi-solid biologically compatible and resorbable matrix) at or

near the site where the effect of the compound is desired. Therapeutic doses are determined to be appropriate by one skilled in the art, see supra.

Furthermore, the present invention relate to the use of

(a) a prokaryotic protein having nucleoside or nucleotide kinase activity for a nucleoside or nucleotide analog or a polynucleotide encoding and capable of expressing said protein *in vivo* or a vector containing said polynucleotide,

(b) the polypeptide, the polynucleotide or the vector of the invention; and/or

(c) the nucleoside or nucleotide analog identified in the method of the invention

for the preparation of a pharmaceutical composition for the activation of nucleoside or nucleotide analogs or nucleoside or nucleotide based prodrugs and/or for the treatment of viral infections and/or diseases or cancer. Preferably said activation results in a cytotoxic nucleoside or nucleotide. In a preferred embodiment of the use of the invention said viral infection is HIV infection.

In a further embodiment the present invention relates to the use of the inhibitor obtainable by the method of the invention for the preparation of a pharmaceutical composition for inhibiting virus replication or for treating cancer. As mentioned above, the inhibitor identified and obtainable by the method of the invention may be used to specifically inhibit the activity of a viral kinase. This would mean that the activation of nucleosides and/or nucleotides necessary for viral replication is suppressed and therefore may result in preventing of the production of viral progeny.

In a preferred embodiment of the invention, the pharmaceutical compositions, e.g., to be used as described above further comprise or are designed to be administered with a nucleoside or nucleotide analog, preferably AZT or d4T.

In a still further embodiment, the present invention relates to a method for the preparative synthesis of a nucleoside phosphate analog comprising:

(a) using a polynucleotide of the invention or as defined above in a noncellular system or in a cell ex vivo, and

(b) formulating the cells modified in step (a) in a pharmaceutically acceptable form.

In another embodiment the present invention relates to the use of

(a) a prokaryotic protein having nucleoside or nucleotide kinase activity for a nucleoside or nucleotide analog or a polynucleotide encoding and capable of expressing said protein *in vivo* or a vector containing said polynucleotide, or

(b) the polypeptide, the polynucleotide or the vector of invention

for the preparation of nucleoside phosphates or analogs and derivatives thereof. Prior to the present inventions it was cost extensive or even impossible to produce analogs of nucleosides or nucleotides useful in diagnostic and therapeutic approaches; examples are phosphorylated forms of AZT and d4T. With the help of the present invention it is now possible to prepare such nucleoside and nucleotide analogs, for example, β or γ-phosphate labeled nucleotides. For instance, the polypeptide of the invention is particularly suited for use in the preparation of nucleoside or nucleotide analogs in which it can be utilized in liquid phase or bound to a solid phase carrier. The polypeptide of the invention can be bound to many different carriers and used to produce nucleoside and nucleotide analogs. Examples of well-known carriers include glass, polystyrene, polyvinyl chloride, polypropylene, polyethylene, polycarbonate, dextran, nylon, amyloses, natural and modified celluloses, polyacrylamides, agaroses, and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding the polypeptide or will be able to ascertain such, using routine experimentation. There are many different labels known to those of ordinary skill in the art. Examples of the types of labels which can be used in the present invention include radioisotopes, colloidal metals, fluorescent compounds, chemiluminescent compounds, and bioluminescent compounds. On the other hand, a cell of the invention may be

used in a fermentation process for the production of such nucleoside or nucleotide analogs.

These and other embodiments are disclosed and encompassed by the description and examples of the present invention. Further literature concerning any one of the methods, uses and compounds to be employed in accordance with the present invention may be retrieved from public libraries and databases, using for example electronic devices. For example the public database "Medline" may be utilized which is available on the Internet, for example under http://www.ncbi.nlm.nih.gov/PubMed/medline.html. Further databases and addresses, such as http://www.ncbi.nlm.nih.gov/, http://www.infobiogen.fr/, http://www.fmi.ch/biology/research_tools.html, http://www.tigr.org/, are known to the person skilled in the art and can also be obtained using, e.g., http://www.lycos.com. An overview of patent information in biotechnology and a survey of relevant sources of patent information useful for retrospective searching and for current awareness is given in Berks, TIBTECH 12 (1994), 352-364.

The pharmaceutical compositions, uses, methods of the invention can be used advantageously for the treatment of all kinds of diseases hitherto unknown as being related to or dependent on viral diseases or cancer. The pharmaceutical compositions, methods and uses of the present invention may be desirably employed in humans, although animal treatment is also encompassed by the methods and uses described herein.

The figures show:

Figure 1 The TP5A-TmpK complex structure. **a.** A ribbon diagram of a monomer (TmpK is a homodimer) with the helices drawn in red and strands in green. The TP_5A bound at the active site is in cyan with the 5 phosphorous atoms in purple. The LID region adopts a helical conformation in the TP_5A bound complex in contrast to a coil with no

secondary structure when only dTMP present. **b.** Stereo view of the TP_5A-TmpK crystal packing. The monomers can be designated active (green) or inactive (red) according to the presence of a symmetry related arginine (depicted as ball-and-stick) in the active site. The TP_5A molecules are shown in cyan. Note that the arginines in green penetrate the active site of the red monomers (unlike the red-colored arginines that are further away) and thus hinder the formation of an active conformation by those monomers.

Figure 2 **a.** Stereo view of a simulated-annealing omit map of TP_5A and arginine 15 from an active monomer contoured at 3 sigma. Arginine 15, which is located in the P-loop (shown as coil running left-to-right) interacts with the middle phosphate of TP_5A. **b.** Stereo view of the active site. The TP_5A nucleotide is drawn in yellow except for the 5 phosphorous atoms that are in purple. **c.** A distance map with the same view as above. P-loop residues are additionally marked with a asterisk.

Figure 3 Steady-state kinetics of dTMP versus AZT-MP with *E. coli* TmpK.

a. The catalytic activity of *E. coli* TmpK (16.4 nM) with dTMP and ATP was measured as described in Methods at dTMP concentrations indicated at the X-axis, and the following ATP concentrations 15 µM (○), 25 µM (●), 50 µM (□), 100 µM (■), 300 µM (△) and 1000 µM (▲). The maximal catalytic activity attained was kcat = 15 s-1 with KM dTMP of 2.7 µM with ATP and dTMP binding with a factor for synergism of 6.

b. The catalytic activity of *E. coli* TmpK (15.1 nM) with AZT-MP and ATP. The ATP concentrations were as indicated in a. The turnover number (kcat) reached a maximum of 6 s^{-1} at saturating concentrations of both substrates. The binding of dTMP and ATP is

synergistic by a factor of 6. The apparent KM of dTMP is lowered from 17 µM (extrapolated to no ATP) to 2.7 µM at saturating ATP concentration, and likewise the apparent KM value of ATP is lowered from approximately 50 µM to 8 µM. This synergism of binding is completely absent with ATP and AZT-MP, the KM value of ATP being around 50 µM independent of the AZT-MP concentration and that of AZT-MP (30 µM) being very close to dTMP (17 µM) in the absence of ATP. Therefore, it is predicted that the 3'-OH group of dTMP may be directly involved in the synergism of ATP and dTMP binding.

Figure 4 Sequence alignment (using GCG) of 3 eukaryotic (KTHY_SCHPO is TmpK from Schizosaccharomyces pombe) and the *E. coli* thymidylate kinase amino-acid sequences. Shaded black are residues conserved in all sequences, in gray are similar amino-acids found in at least 3 of the sequences. The secondary structural elements (helices as tubes, strands as arrows) of the yeast TmpK are also shown. Residues that contact TP_5A directly are marked with an arrow. The LID part of the sequence is marked as a dotted line; note the difference in the P-loop region between the eukaryotic TmpKs and the *E. coli* enzyme (between β1 and α1), and the additional basic residues found only in *E. coli* TmpK.

Figure 5 Comparison of yeast thymidylate kinase (thick line) with HSV1 thymidine kinase (thin line) (pdb code 1kin). **a.** Superposition of both kinases (both are homodimers; only a monomer is shown) with TmpK in blue and HSV1-TK in yellow. The TP_5A bound to TmpK is also shown (cyan) as is the thymidine bound to HSV1-TK (orange). **b.** Stereo view (rotated 180° relative to the figure above) with a close-up of the active sites. Thymidine (orange) is displaced relative to the thymidine moiety of TP_5A (cyan). As a result, where in TmpK Tyr102 discriminates against ribonucleotides, the corresponding Tyr172 in

HSV1-TK makes a base-stacking interaction with the thymidine base, a role accomplished by Phe69 in TmpK.

Abbreviations:

TmpK, thymidylate kinase
UmpK, uridylate kinase
AK, adenylate kinase
HSV1-TK, herpes simplex virus thymidine kinase, type I
AZT, 3'-deoxy-3'-azido thymidine
AZT-MP, 3'-deoxy-3'-azido thymidine monophosphate
AZT-DP, 3'-deoxy-3'-azido thymidine diphosphate
AZT-TP, 3'-deoxy-3'-azido thymidine triphosphate
TP_5A, P1-(5'-adenosyl) P5-(5'-thymidyl) pentaphosphate
UP_5A, P1-(5'-adenosyl) P5-(5'-uridyl) pentaphosphate
PA, PB, PC, PD, PE, phosphate groups of TP_5A
NMP, nucleoside monophosphate kinases
ncs, non-crystallographic symmetry
RMSD, root mean square deviation

The examples illustrate the invention.

Thymidylate kinase (E.C. 2.7.4.9; ATP:dTMP phosphotransferase) catalyzes the phosphorylation of thymidine monophosphate (dTMP) to thymidine diphosphate (dTDP) utilizing ATP as its preferred phosphoryl donor (Jong & Campbell, 1984) according to the scheme:

$$dTMP + ATP\ \ Mg \Leftrightarrow dTDP + ADP \bullet Mg$$

Its location at the junction of the de-novo and salvage pathways for thymidine triphosphate (dTTP) synthesis makes thymidylate kinase (TmpK) an essential

enzyme for cell proliferation, and thus an attractive target for the development of drugs against cancer. In addition to its physiological role, TmpK is also involved in the activation of the AIDS drug 3'-azido-3'-deoxythymidine (AZT). AZT is a prodrug that must be phosphorylated three times to its triphosphate form (AZT-TP), since it is AZT-TP that inhibits viral replication by DNA chain termination. TmpK, which catalyzes the second phosphorylation step, from the monophosphate (AZT-MP) to the diphosphate (AZT-DP), has been shown to be the rate limiting enzyme in the AZT activation pathway (Furman et al., 1986, Qian et al., 1994). This results in a toxic accumulation of AZT-MP to millimolar concentration in cells exposed to AZT (Bridges et al., 1993, Tornevik et al., 1995, Yan et al., 1995), and in a low concentration of the active compound AZT-TP. Slow activation rates of prodrugs have been implicated in allowing the replicating virus to select for resistant mutants.

The Examples of the present invention provide the understanding the mechanism of phosphoryl transfer and pinpoint the amino acid residues involved and thus provide for a generally applicable method for developing and improving strategies for the treatment of cancer and AIDS. For example, the design of a mechanism-based inhibitor of TmpK resulting in the halt of dTTP synthesis, and thus cell proliferation, could play a role in chemotherapy of cancers.

Example 1: Structure of TmpK from yeast

TmpK from *Saccharomyces cerevisiae* (cdc8 gene), has 216 amino acid residues and a molecular weight of 25 kD Crystals of the complex between the bisubstrate inhibitor TP_5A with yeast thymidylate kinase were obtained by the hanging drop method. A 12 mg/mL enzyme solution was premixed with a TP_5A solution to a final concentration of 10 mg/mL enzyme and 2 mM TP_5A. Equal volumes of the protein-nucleotide solution and a solution composed of 20% PEG 2000 monomethyl ether, 100 mM sodium acetate pH 4.6, and 200 mM ammonium sulfate were mixed, and left to equilibrate at room temperature against a reservoir composed of the latter solution. Crystals with typical dimensions (μm) of 400 × 200 × 100 grew within

days in space groups P1, $P2_1$, or $P2_12_12$ with very similar unit cell dimensions. This space group polymorphism made these crystals unsuitable for the initial structure determination (Lavie et al., 1997b). The structure reported here was solved by molecular replacement using the dTMP-TmpK complex structure as starting model (see below). A data set was collected at 100 K from a crystal soaked shortly in mother liquor with 25% glycerol as cryo-protectant, using a Siemens multiwire area detector mounted on a Mac Science rotating anode operating at 45kV, 100 mA. Processing of the data was carried out with XDS (Kabsch, 1993). The structure reported here is from a crystal which reduced in space group $P2_1$ and diffracted to 1.8 Å resolution, see Table 1.

Table 1: Data Collection and Refinement Statistics

Data collection	
Temperature (K)	100
Resolution range (Å)	43.8 - 1.8
Observed reflections	270178
Unique reflections	116451
Completeness (%, overall/last shell)	82.0/56.9
R_{sym}[a] (%, overall/last shell)	5.9/26.6
Space group	$P2_1$
Unit-cell (Å, °)	a=72.6 b=87.3 c= 155.0 β=90.1
molecules / asymmetric unit	8
Refinement Statistics	
Resolution range (Å)	43.8 - 2.0
R_{factor}[b]/R_{free} (%)	20.9/27.9
RMS deviations	
bond lengths (Å)	0.012
bond angles (Å)	1.64
dihedral angles (°)	24.7
improper angles(°)	1.34
Reflections with F>0 s (working / test)	101931 / 5401

# of protein atoms	13791
nucleotide atoms	440
water molecules	1198
Average B (Å^2)	
main chain	20.5
side chain	22.1
waters	28.6
nucleotides	16.3

[a] R_{sym}=_|I-<I>|/_I
[b] R_{factor}=_||F_{obs}| - |F_{calc}||/_|F_{obs}| 5% of reflections were used for R_{free}.

The TP_5A complex structure (Figure 1) was solved by the molecular replacement method with the partially refined dimeric dTMP complex structure as the starting search model. As mentioned above, the TP_5A co-crystals exhibit space group polymorphism; the data set used has $P2_1$ symmetry with 4 dimers in the asymmetric unit. The pseudo-symmetry with the orthorhombic space group suggested the presence of two tetramers in the monoclinic asymmetric unit. This implied that when searching with the TmpK dimer, one should expect 2 rotation function solutions, where each solution has two different translation function solutions, thus yielding the two tetramers. Both AMoRe (Navaza, 1994) and X-PLOR (Brünger, 1993) yielded the two expected rotation functions solutions, but it was not possible to find the two appropriate translation function solutions. The self-Patterson map showed one clear peak that was interpreted as the vector relating the two tetramers. Thus, each of the two rotation function results were first applied on the search dimer, and then the translation vector from the self-Patterson was applied on the rotated dimers. The resulting two pairs of tetramers were used independently as search models for the translation function and yielded clear solutions. In $P2_1$ the origin of the unique axis is not defined so the relative position of one tetramer to the other was found by arbitrarily keeping one tetramer at y=0, which also limits the translational search to two dimensions (x and z). At first, the refinement was carried out using strict non-crystallographic symmetry (ncs) thus allowing to work with a dimer instead of the octamer. The refinement proceeded

smoothly with clear electron density for the TP_5A bisubstrate inhibitor. When the R-factor reached 27% the ncs constraints were totally relaxed. The R-factor converged at 20% and 29% for the work and test set (5% of total reflections), respectively. Ncs restraints were reapplied at relatively low weight causing a slight decrease in R_{free} and a bigger increase in R_{work}. Consequently, the final model was refined with ncs-restraints grouping monomers 1, 3, 5, and 7 in one group and 2, 4, 6, and 8 in the other (the LID sequence and nucleotides were not included). The numbering used is residues 1 to 216 and 501 to 716 for monomer 1 and 2, respectively, with TP_5A numbered 217 in monomer 1 and 717 in monomer 2. Monomers 3 and 4 are numbered as above with the addition of 1000, 5 and 6 with the addition of 2000, and 7 and 8 with the addition of 3000 (thus TP_5A 3217 is the TP5A bound to monomer 7). Model building was done with the program O (Jones et al., 1991). The present model consists of 8 monomers with some residues omitted because of poor electron density (residues 1, 2, 1137 to 1148, 2001, 3001, 3501, 3502) and some residues modeled as alanines, 8 TP_5A molecules, and 1198 water molecules (see Table 1). A simulated-annealing omit map (calculated without the TP_5A and Arg15) with the current TP_5A model superimposed is shown in Figure 2*a*.

Unlike the known structures of other nucleoside monophosphate (NMP) kinases (adenylate kinase, uridylate kinase, guanylate kinase) which are all monomeric, TmpK is a homodimer. Despite no amino acid sequence similarity between TmpK and any of the other NMP kinases, TmpK assumes the common fold of the other NMP kinases with 5 parallel beta strands forming a beta-sheet core surrounded by helices (Figure 1*a*). The highly hydrophobic dimer interface is composed of 3 parallel a-helices provided by each monomer that stack against each other.

The asymmetric unit of the $P2_1$ crystal form consists of eight monomers which pack in 4 dimers. A homodimer is the basic unit of yeast TmpK, but no cooperativity was detected in our kinetic assays (Lavie et al., 1997a). Upon overlaying all 8 monomers, it was evident that they can be divided into two groups; mon1, mon3, mon5 and mon7 adopt one conformation, whereas mon2, mon4,

mon6, and mon8 adopt another conformation (as neither ncs constraints nor restraints where used in the last stages of refinement, the conformation adopted by each monomer is totally independent of the other monomers; in the final cycle of refinement ncs-restraints were reapplied). For reasons that become apparent later, the first group was designated as inactive monomers, and the second group of monomers as active (Figure 1*b*). The RMSD between the active and inactive monomers excluding the LID sequence is 0.3 Å (on Ca atoms).

Most likely it is the crystal packing which determines which monomers adopt the active conformation, and which do not. Apparently, an active monomer of one dimer inactivates a monomer of another dimer. Arg173 (a non-conserved residue) from all active monomers interacts with the TP5A bound to the inactive monomers through a oxygen atom of phosphate PB, whereas Arg173 from the inactive monomers cannot form such an interaction. The presence of Arg173 prevents the formation of an active conformation of those monomers. Thus, it is this interaction of Arg173 from an active monomer with the TP5A of an inactive monomer which presumably prevents the closure of the LID of those monomers and results in poor density. Since TmpK is a dimer in solution (based on gel-filtration and dynamic light scattering experiments) and the fact that binding studies with TP5A clearly indicate a single class of binding sites in solution, it is assumed that the partitioning into active and inactive monomers is a pure crystallization artifact with no physiological significance. Therefore, the discussion is limited to the active conformers.

Example 2: Comparison with dTMP-TmpK complex structure

Kinases undergo large conformational changes upon the binding of substrates (Vonrhein et al., 1995). At least 3 different states have been described; an open state in the absence of substrates, a partially closed state when one of the substrates is bound, and a closed state when both substrates are present. The previously reported dTMP-TmpK complex (Lavie et al., 1997b) represents the partially closed state (ATP is missing) whereas the TP5A complex represents the

fully closed state, which is catalytically competent. Superposition of these two complexes shows that the main difference between them lies in the LID region (the other slight differences concentrate in loop sequences, which are expected to vary as a result of the different crystal packing between the two complexes). In NMP kinases the LID region has been observed to change its conformation upon ATP binding. In the dTMP-complex the LID sequence was only traceable in one of the monomers, and the density for the traced sequence was weak, indicating disorder, as expected in the absence of ATP. With TP_5A it was expected to see a much more ordered LID region. In the monomers that were designated as the inactive conformer, the LID region is very hard to trace and lacks secondary structure. In contrast, the LID region of the active conformers was easier to trace and forms a helix (Figure 1*a*). Upon superposition of TmpK with uridylate kinase (see more later for a detailed comparison with other NMP kinases), this newly formed helix overlays very well with a helix from the UmpK LID region (Scheffzek et al., 1996).

Example 3: Binding of TP_5A

Titration of TmpK to a cuvette containing 0.09 mM of the fluorescent bisubstrate inhibitor TP_5A-MANT results in a fluorescence increase that indicates complex formation. Equilibrium fluorescence measurements were performed as described (Reinstein et al., 1990) using a SLM 8100 spectrofluorimeter with an excitation wavelength of 360 nm and emission wavelength of 440 nm. The experiments were done at pH 7.5 in a solution containing 50 mM Tris/HCl, 5 mM $MgCl_2$, 2 mM EDTA, and 100 mM KCl with the temperature of the cuvette being held constant at 25 °C. The fluorescent N-methylanthraniloyl group (MANT) joined to the ribose of the adenosine moiety of the bisubstrate inhibitor TP_5A (TP_5A-MANT) was used as probe (Reinstein et al., 1990).

Time resolved binding studies were performed with the same buffer as described above with a Hi-Tech Scientific SF61 stopped flow apparatus. As a signal for complex formation either the extrinsic signal of TP_5A-MANT with excitation at 360 nm and a cutoff at 420 nm or the intrinsic tryptophan signal with excitation at 295

nm and a cutoff at 320 nm was used. These experiments were performed essentially as described (Packschies et al., 1997). The titration data were fitted by a quadratic equation (Reinstein et al., 1990) which yields a K_d of 135 nM for the binding of TP_5A-MANT to TmpK. Competitive displacement of TP_5A-MANT from the so formed complex by TP_5A and analysis by a cubic equation (Thrall et al., 1996) shows the K_d for the binding of TP_5A to TmpK to be 95 nM. The affinity of TP_5A to TmpK appears to be rather weak compared to *E. coli* adenylate kinase that binds AP_5A with a K_d of 15 nM or to *D. discoideum* UmpK that binds UP_5A with a K_d of 3 nM (Reinstein et al., 1990, Wiesmüller et al., 1995). Time resolved measurements show that the binding of TP_5A and TP_5A-MANT is relatively fast but perhaps somewhat below the diffusion controlled limit with $k_{on} = 10 \times 10^6\ M^{-1}\ s^{-1}$ and $k_{on} = 5.5 \times 10^6\ M^{-1}\ s^{-1}$, respectively. This indicates that the conformational changes induced by the bisubstrate inhibitor, e.g. closing of the LID region over substrate, do not constitute the rate limiting step of the catalytic cycle since k_{cat} of yeast TmpK was shown to be 35 s^{-1} (Lavie et al., 1997a) whereas the binding of TP_5A did not deviate from a linear relationship for k_{obs} versus [TP_5A] up to 60 s^{-1}.

In the crystal the bisubstrate inhibitor TP_5A appears to be bound tightly by all monomers (see Figure 2), as can be inferred from the relatively low B-factors (see Table 1) and the excellent electron density for the nucleotides. Before describing in detail the interactions made between the TP_5A di-nucleotide and the enzyme, it is important to realize that the TP_5A molecule is not a transition-state analog but is rather a bisubstrate or a biproduct analog (Scheffzek et al., 1996). In other words, it is not clear whether the observed TP_5A is closer to the ATP-dTMP or the ADP-dTDP state.

The thymidine part of TP_5A, which is completely surrounded by protein atoms, has a lower B-factor than the relatively exposed adenine part. The detailed interactions between the thymidine moiety and TmpK has been already described for the dTMP-TmpK complex (Lavie et al., 1997b), so only the differences in the TP_5A complex will be outlined here. Lysine 37 is observed to interact with the phosphate of dTMP in the dTMP-TmpK complex structure. This phosphate (for phosphate

notation see figure 2*c*) corresponds to PE in TP5A. In the TP5A complex Lys37 interacts with PD instead; PD can be seen either as the connector between dTMP (PE) and ATP (PA, PB, PC), and thus as an artificial moiety, or as the b phosphate of dTDP. In the other NMP kinase structures with such bisubstrate analogs, the position of this connecting phosphate varies (Abele & Schulz, 1995, Müller & Schulz, 1992, Scheffzek et al., 1996). Thus, it is hard to interpret this change in Lys37 interaction from PE to PD, but it is possible that this residue participates in stabilizing the transferred phosphate group in the transition state.

The P-loop motif has been identified in many ATP- and GTP-binding proteins (Saraste et al., 1990). The binding of the nucleotide to the P-loop is through main-chain nitrogen atoms to the a and b phosphates of the nucleotide, and through a strictly conserved lysine. The ATP part of TP5A interacts in such a manner with the P-loop of TmpK, through the amides of Thr16 and Thr19 to the PB oxygen atoms, and Thr20 to PA oxygen atoms. In addition to the amide-phosphate interaction of Thr20, its side-chain also interacts with PA. The P-loop lysine (Lys18) is observed to interact with the PB and PC oxygen atoms of TP5A. It is positioned such that it may stabilize the transition state, in agreement with mutational analysis of adenylate kinase (Reinstein et al., 1988, Reinstein et al., 1990, Tsai & Yan, 1991). In the non-active monomers, Arg15 (situated at the tip of the P-loop) is in an extended conformation making no interaction with TP5A, while in the active monomers Arg15 bends to interact with a PC oxygen. It is in fact this interaction which lead the inventors dividing the monomers into active and inactive, since the inventors attribute an essential catalytic role to this arginine (see below).

The adenine part of TP5A makes only two interactions with the enzyme, both through the amino group at C6. One is to the main-chain carbonyl group of Lys187, the other through a water molecule to the highly conserved Gln21. This would explain the preference of TmpK for adenine nucleotides over guanine; in guanine-based nucleotides there is a carbonyl group at C6 instead of the amino group, which would prevent the favorable interaction with the main-chain carbonyl of Lys187. However, the partial acceptance of non-adenine-based nucleotides as phosphoryl donors by TmpK (Jong & Campbell, 1984) is consistent with our

structure, as there are no strong interactions between the adenine base and the enzyme, and there would be no steric discrimination against the amino group at the C2 position of guanine, which would face towards the solvent.

Bound magnesium was not observed in the structure, which is attributed to the low pH and the presence of ammonium sulfate in the crystallization conditions. Inferring from other NMP kinases (Scheffzek et al., 1996), the magnesium is expected to be octahedrally coordinated to the PB and PC oxygens (corresponding to b and g phosphates of ATP), Thr16, Asp93 (directly or through a water molecule), and two additional water molecules. At the position where magnesium is expected electron density was seen only for a water molecule bridging Asp93 and PC.

Example 4: Comparison with other NMP kinases

In addition to this TP_5A-TmpK complex structure, there are at present 3 other structures of NMP kinases with bisubstrate analogs; adenylate kinase (yeast and *E.coli*) with AP_5A (Abele & Schulz, 1995, Müller & Schulz, 1992), and uridylate kinase (slime mold) with UP_5A (Scheffzek et al., 1996). The above mentioned adenylate kinases (AK) are both of the long LID region type, while uridylate kinase is more like TmpK having a short LID region (for the LID's amino acid sequence see Figure 4). Therefore, the inventors have chosen to compare TmpK to uridylate kinase (UmpK). The structure of UmpK complexed with UP_5A was superimposed on the TP_5A-TmpK complex structure by aligning the P-loop sequences of both enzymes. While UmpK and AK have a high degree of sequence identity, none exists between TmpK and either UmpK or AK, but structurally, TmpK overlays on UmpK surprisingly well and UP_5A occupies nearly the same position as TP_5A. There is, however, a striking difference. The LID domain of UmpK interacts directly with the phosphates of UP_5A via the basic residues Arg131 and Arg137, which have been shown to stabilize the transition state (Schlichting & Reinstein, 1997). While lacking such basic residues in the LID domain, TmpK has Arg15 in the P-loop, the corresponding amino acid in UmpK being a glycine. And in fact, in what

the inventors defined as the monomers in the active conformation, Arg15 makes a 2.8 Å hydrogen-bond interaction with one oxygen atom of PC. This would be an example of the importance of having catalytic residues in proximity of the reaction center, but their origin, LID or P-loop, is irrelevant for catalysis.

Example 5: Comparison with Herpes Simplex Virus-1 thymidine kinase

From sequence similarity the Herpes Simplex Virus-1 thymidine kinase (HSV1-TK) has been grouped with the thymidylate kinase family. In fact, HSV1-TK has both a thymidine and a thymidylate kinase activity, suggesting that it evolved from thymidylate kinase and acquired the additional thymidine kinase activity. The 3-dimensional X-ray structure of HSV1-TK has been solved recently revealing, as for TmpK, a dimer as the basic unit (Brown et al., 1995). The viral kinase, however, contains 376 amino acids compared to TmpK's 216. It has additional N-terminal (45 residues) and C-terminal sequences (87 residues), as well as a few short inserts in the overlapping sequence region. Nevertheless, the overlapping sequence regions can be relatively easily overlayed structurally, with both kinases displaying the basic kinase fold as previously described (Figure 5).

The low level of substrate specificity of the viral kinase, which is medicinally important in the activation of nucleoside prodrugs, can now be rationalized upon comparison with the highly specific TmpK. Using the 14 amino-acid stretch of the P-loop regions of both enzymes for the superposition (Ca rms of 0.7 Å) it becomes apparent that the base moiety of thymidine bound to HSV1-TK is about 3.5 Å displaced in relation to the position of the base in TmpK. The displacement is away from the core of the enzyme towards the solvent. This slippage of the nucleotide towards the active site residues (mainly basic residues which would interact with phosphate groups) but without the concomitant contraction of the nucleotide's base binding pocket explains why HSV1-TK can phosphorylate even purines in addition to pyrimidines, and how it can possess both thymidine and thymidylate kinase activity. Presumably, after the first phosphorylation of thymidine to thymidine monophosphate, the nucleotide has room to slip deeper into the

binding site, thus placing the phosphate at the position previously occupied by the C5 hydroxyl. Now the second phosphorylation can take place. Thus, it is the nucleotide that moves relative to the enzyme's catalytic residues. This would suggest that HSV1-TK, while accepting purine nucleosides as substrates, might not accept purine monophosphates (the substrates for the thymidylate kinase activity) as they must be bound deeper into the binding cavity and thus might not fit. And in fact, the guanine based anti-herpes drug acyclovir is phosphorylated to the monophosphate by HSV1-TK, but the following phosphorylation step is catalyzed by guanylate kinase (Elion, 1982), consistent with this interpretation.

In contrast to the base, other residues in the active site overlay very well, noteworthy being Tyr102(TmpK) and Tyr172 (HSV1-TK) (within 0.3 Å). Tyr102 is responsible for the discrimination against ribonucleotides in TmpK by being located 3.5Å from the C2 of the ribose. In HSV1-TK it fulfills a completely different role by base stacking with thymine (in TmpK Phe69 stacks analogously to this Tyr with the base but from the opposite side of the base; see Figure 5*b*).

Important for the previous discussion of catalysis are the P-loop and LID regions; in this respect HSV1-TK is more similar to the *E.coli* TmpK than to the yeast TmpK in lacking the arginine in the P-loop (D14R15 in yeast, E12G13 and H58G59 in *E.coli* and HSV1-TK, respectively) and having basic residues in the LID region. Arg222 of HSV1-TK, which is in the LID region, is situated less than 4.5 Å from the deoxyribose O5' of thymidine. Analogously to other NMP kinases that utilize basic residues from the LID region to stabilize the negative charge developing in the transition state, we propose that Arg222 of HSV1-TK fulfills this catalytic role.

Example 6: Implications for catalysis and AZT activation

In addition to phosphorylating dTMP to dTDP, thymidylate kinase is also part of the activation pathway of the anti-HIV prodrug 3'-deoxy-3'-azidothymidine (AZT). In fact, TmpK is the rate-limiting enzyme in this activation pathway. Together with the previously reported nucleotide-complex structures with dTMP and with AZT-MP

(Lavie et al., 1997b), the TP_5A-TmpK complex structure suggests which residues are needed for catalysis, and why AZT-MP is phosphorylated by TmpK so poorly. Whether the phosphoryl transfer mechanism is of an associative nature, a dissociative one, or of a mixed type, the negative charge developing during the transition state compared to the ground state has to be stabilized. This would be achieved by a basic residue (Arg or Lys), and its interaction with the phosphate must be made or strenghtened at the transition state. In UmpK it has been shown that the mechanism of phosphoryl transfer is most probably associative, and that the charge developing on the transferred phosphoryl group is stabilized mainly by Arg131 and Arg137 from the LID region, in addition to Lys19, Arg148 and the catalytic magnesium ion. In contrast, yeast TmpK lacks basic residues in its LID region, but does have an arginine residue at the tip of the P-loop. In the TmpK-TP_5A complex this arginine's side-chain is at a similar position as would be an arginine originating from the LID domain, and is located 2.8 Å from PC of TP_5A. Therefore, the inventors attribute to Arg15 a similar role to that of the UmpK's LID arginines.

As pointed out previously (Lavie et al., 1997a, Lavie et al., 1997b), the binding of AZT-MP causes the P-loop of TmpK to shift by about 0.5 Å due to the interaction between the azido moiety of AZT and the side-chain of Asp14. This shift affects Arg15 as well, causing it to be not optimally located to fulfill its catalytic role. Most, but not all thymidylate kinases sequenced to date have an arginine in the P-loop. A noticeable exception is thymidylate kinase from *E. coli* which has a glycine residue at that position (like UmpK). The LID domain of the *E. coli* TmpK contains basic residues, unlike yeast TmpK but like UmpK. The present inventors postulate that the *E. coli* TmpK functions very similarly to the slime mold UmpK, where arginine residues from the LID domain participate in catalysis (candidates are Arg147, Arg149, Arg151, and Arg156). This would suggest that a P-loop movement of *E. coli* TmpK, like that observed with the yeast enzyme and AZT-MP, would not have such a detrimental effect on catalysis, as the residues important for catalysis (except the conserved lysine) do not originate from the P-loop. To test this

hypothesis, thymidylate kinase from *E. coli* was cloned, expressed and purified and its dTMP and AZT-MP phosphorylation rates were compared.

Example 7: Steady-state kinetics of *E. coli* TmpK

As predicted based on the structural rational explained above, the rate of phosphorylation of AZT-MP and dTMP by *E. coli* TmpK is comparable (Figure 3). In steady-state kinetics assays the catalytic activity of TmpK was measured with a coupled colorimetric assay essentially as described (Reinstein et al., 1988) with the following assay buffer: 100 mM Tris/HCl pH 7.5, 200 µM NADH, 400 µM phosphoenolpyruvate, 80 mM KCl, and 1.4 mM $MgCl_2$ at 25 °C. In addition, the assay contained different concentrations of ATP and dTMP and 10 to 50 nM TmpK as indicated in Figure 3. The data were analyzed with the Michaelis-Menten equation and the non-linear regression program Grafit (Erithacus software). While in the case of yeast TmpK AZT-MP is phosphorylated 200 fold slower than dTMP, for the *E. coli* enzyme the factor is only 2.5 (Table 2),

Table 2: Steady-state kinetic parameters

	yeast	*E. coli*
K_M for dTMP with ATP	9 µM	2.7 µM
K_M for AZTMP with ATP	6 µM	30 µM
K_M for ATP with dTMP	190 µM	8 µM
K_M for ATP with AZTMP	300 µM	50 µM
k_{cat} with ATP and dTMP	35 s^{-1}	15 s^{-1}
k_{cat} with ATP and AZTMP	0.175 s^{-1}	6 s^{-1}
ratio k_{cat} dTMP/AZTMP	200	2.5
ratio k_{cat}/K_M dTMP/AZTMP	133	27.5

illustrating the big difference in acceptance of AZT-MP by these two dTMP-kinases from different organisms. It appears from this comparison that NMP kinases have developed different strategies to attain catalysis; one class has catalytic residues in the P-Loop and the LID regions (AK, UmpK, etc.) whereas most dTMP kinases (*E. coli* TmpK being an exception) apparently lack the catalytic arginine residues of the LID region. In this respect it has also to be considered that the mechanism of phosphoryl transfer of the yeast TmpK may not be purely associative as was deduced for UmpK (Schlichting & Reinstein, 1997).

Example 8: Substitution of Argenine 15 with glycine in the P-loop of yeast TmpK and replacement of the yeast LID-region with the *E. coli* LID-region

In order to prove that the observations described in examples 1 to 7, supra, can be indeed applied to a nucleoside or nucleotide kinase naturally having only a low catalytic activity for nucleoside or nucleotide analogs, the yeast TmpK enzyme described above has been mutated at position X_3 in the consensus sequence of the P-loop (amino acid position 15 in the amino acid sequence of the yeast TmpK enzyme; see Figure 4) in order to substitute the amino acid argenine with glycine, the amino acid at the corresponding position in the P-loop of *E. coli* TmpK; see Figure 4. Furthermore, the LID sequence of the yeast TmpK enzyme from position 131 to 150 (FLSTQ ... GDER) was replaced by the corresponding *E. coli* TmpK LID sequence from position 138 to 150 (YLDVTP ... ELDR); see Figure 4. The mutations referred to above were introduced into the yeast TmpK either alone or in combination by introducing the corresponding modifications in the DNA sequence underlying the amino acid sequence of the yeast TmpK as described in Sambrook et al., supra. The resultant enzymes were then tested and compared with respect to dTMP and AZT-MP phosphorylation rates as described in Example 7 and in Reinstein et al., (1988) with 10 mM $MgCl_2$ and the following, constant concentration of nucleotides: 2 mM ATP and either 1 mM TMP or 1mM AZT-MP. Furthermore, as a control the TmpK enzymes from yeast, *E. coli*, human, mouse and herpes simplex virus were tested. The results are shown in Table 3 below.

Tabl 3

Protein	Activity TMP	Activity AZT-MP	Reduction
yeast	100	0.5	200
E. coli	43	21	2
yeast R → G	0.5	-	-
yeast → LID *E. coli*	4	0	-
yeast R → G + → LID *E. coli*	25	2	12
Human	5	0.07	70
Mouse	20	0.8	25
Herpes S.	1	0.5	2

As can be inferred from Table 3, the mutated version of the yeast TmpK enzyme that contains both, the mutated P-loop and LID-region exhibits improved kinase activity for the nucleotide analog AZT-MP which is 30-fold higher than that of the corresponding human enzyme. Moreover, the activity of the mutated yeast TmpK exceeds that from mouse and herpes simplex virus, the latter of which has been speculated to be useful in gene therapy.

Example 9: Improvement of kinase activity of human thymidylate kinase

Modifications at two different positions of a nucleoside monophosphate kinase sequence are listed as the means of achieving a kinase with enhanced prodrug phosphorylation activity. Those positions are either the P-loop region, the LID region, or both. The human thymidylate kinase (hTMPK) has been modified accordingly with the result that higher azidothymidine monophosphate (AZTMP) phosphorylation activity was achieved. In addition, a modification outside of the P-loop or the LID region, namely that of phenylalanine 105 to tyrosine, also results in a hTMPK variant having the desired kinetic activity with AZTMP. The mutations in the amino acid sequence of hTMPK (SEQ ID NO: 5) and testing of the resultant

polypeptides were performed as described in Example 8. The results are summarized in Table 4.

Table 4

Protein	Activity TMP	Activity AZTMP	ratio TMP/AZTMP
yeast TMPK (WT†)	100‡	0.5	200
A) human TMPK (WT)	1.9	0.019	100
B) Arg16Gly Small *E.coli* LID§	1.9	0.64	33
C) Arg16Gly Large *E. coli* LID¶	4.3	6.1	0.7
D) Phe105Tyr	0.475	0.665	0.7

†WT: wild-type

‡The rate of yeast TMPK with TMP is set at 100.

§Small *E. coli* LID: exchange of hTMPK LID sequence with that of E. coli, residues 145-148.

¶Large *E. coli* LID: exchange of hTMPK LID sequence with that of E. coli, residues 135-148.

The modifications of hTMPK in the P-loop (Arg16Gly) and in the LID region (large *E. coli* LID) have resulted in a variant (marked (C) in the table above) with 320 fold higher activity with AZTMP as substrate in comparison to the wild type hTMPK. Moreover, the substrate specificity has been changed dramatically, such that variant (C) phosphorylates AZTMP at a faster rate than the physiological substrate

TMP. This should improve its efficacy even further since TTP and AZTTP compete for incorporation into nascent viral DNA chains (e.g. TTP and AZTTP compete as substrates for reverse transcriptase). Variant (D), the result of a single amino-acid change, has likewise a faster rate with AZTMP than with TMP, but is about 10 times slower than variant (C).

The results obtained in accordance with the present invention are of far reaching medicinal importance since it for the first time enables a rational approach to using gene therapy to improve the antiviral impact of nucleoside and nucleotide analogs, such as AZT as outlined previously (Balzarini et al., 1988, Guettari et al., 1997, Lavie et al., 1997a).

R ferences:

Abele, Protein Sci 4 (1995), 1262-71.
Balzarini, Biochem. Pharmacol. 37 (1988), 897-903.
Bridges, Biochem Pharmacol 45 (1993), 1571-6.
Brown, Nature Structural Biology 2 (1995), 876-81.
Brünger, X-PLOR: a system for X-ray crystallography and NMR (1993), Yale University Press, New Haven, CT.
Elion, American Journal of Medicine 73(1A) (1982), 7-13.
Furman, Proc Natl Acad Sci U S A 83 (1986), 8333-7.
Jones, Acta. Cryst. A 47 (1991), 110-119.
Jong, J Biol Chem 259 (1984), 14394-8.
Kabsch, J. Appl. Cryst. 24 (1993), 795-800.
Lavie, Nature Medicine 3 (1997a), 922-924.
Lavie, Nature Structural Biology 4 (1997b), 601-604.
Müller, J Mol Biol 224 (1992), 159-77.
Navaza, Acta Cryst. A 50 (1994), 157-163.
Packschies, Biochemistry 36 (1997), 3417-22.
Qian, Biochem Pharmacol 48 (1994), 383-9.
Reinstein, Biochemistry 27 (1988), 4712-20.
Reinstein, Biochemistry 29 (1990), 7440-50.
Reynes, J Bacteriol 178 (1996), 2804-12.
Saraste, Trends Biochem Sci 15 (1990), 430-4.
Scheffzek Biochemistry 35 1996), 9716-27.
Schlichting, Biochemistry (1997).
Studier Methods Enzymol 185 (1990), 60-89.
Thrall, Biochemistry 35 (1996), 4609-18.
Tornevik, Biochem Pharmacol 49 (1995), 829-37.
Tsai, Biochemistry 30 (1991), 6806-18.
Vonrhein, Structure 3 (1995), 483-90.
Wiesmüller, Febs Lett 363 (1995), 22-4.
Yan, J Biol Chem 270 (1995), 22836-41.